IMPROVING SIGNAL TO NOISE RATIO IN CHROMATOGRAPHY

INVENTOR
Robert W. Allington

The present application is a continuation in part of U.S. Patent Applications Serial Number 10/410,373, filed on 09 April, 2003 and Serial Number 10/636,153 filed on 07 August 2003, both of which are herein incorporated by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to a method and apparatus for increasing the signal to noise ratio of a signal received from a photocell used in capillary High Performance Liquid Chromatography (HPLC), Capillary Electrophoresis (CE) or Capillary Electroendoosmosis Chromatography (CEC). For instance, in capillary HPLC a capillary tube serves as the chromatographic column. If the capillary is an open tube, its inside diameter may be from 10μm (detector sensitivity) to 100μm (detector volume). Packed capillary columns have analogous volume limits.

More particularly the present invention makes use of a plurality of photocells. In one embodiment, the photocells are used to detect a solvent spike (absorbance spike or refractive index spike) of a quiescent fluid. In another embodiment, a set of photocells receive light through the same particle of fluid by accurately tracking the solvent spike as it passes the linear array of photocells. In either case, the signals created by all the photocells are summed or integrated over time to increase the signal to noise ratio. In still another embodiment, a noisy signal is divided up into segments. Within each segment, the signal is summed or integrated. The resulting segments are concatenated to produce a new signal having less noise than the original. This invention applies to HPLC, CE, and CEC. References herein to capillary

chromatography or capillary HPLC may also apply to CEC and CE. "Solvent spike" in the claims and elsewhere includes both relating to an absorbance spike as well as refractive index spikes.

Because the present invention is applicable to HPLC, CE, and CEC, these three will be referred to collectively hereinafter as "capillary separation schemes." The term "separation schemes" will encompass all the above, plus those used with small volume columns.

Background Art

To identify components and their concentration in a mixture of solute and solvent using high performance liquid chromatography, light is passed through the solution. The light that is neither reflected nor absorbed impinges on a photocell, where the intensity of the light is converted to an electrical signal. The intensity of the light hitting the photocell is related to the concentration of a particular solute in the solution. Sensitivity of this system is proportional to the path length of the light as it passes through the sample. Recall that the solution is contained in a capillary tube. Increasing the optical path length in the "usual fashion," however, invites problems related to widening and "blurring" of peaks because of Poiseulle flow distribution or worse, separation, recirculation, or a flow with a helical path, along the light path.

Sources of noise in the signal include the light source, thermal effects, and turbulence in the flow of the solution. The signal to noise ratio of the signal produced by a single, stationary photocell may be too low to be useful. It becomes difficult or impossible to pick out peaks in the signal because the signal is extraordinarily low and the noise level is as high as in a larger diameter (e.g. 4.7 mm i.d.) chromatographic flow system. A signal to noise ratio of about 2 is considered the lowest acceptable. To overcome this difficulty, it is possible to move the light source and photocell along a capillary tube through which the solute is flowing; or to move the capillary tube past the light source and sensor. The relative velocity at which the capillary tube moves compared to the light source and photocell is equal to the maximum velocity (in any infinitesimally thin slice of fluid in cross-section) of the

fluid. The purpose is to generate a "signature" of a particular particle of fluid moving at the centerline speed of the flow. This approach has its weaknesses, including the need for accurately correlating the relative positions of the detector and capillary tube with the peaks observed. The need for moving parts or specially triggered and secondarily detected flash from a flashlamp increases the complexity of the apparatus and the potential for failure.

R. E. McKean, in his University of Massachusetts Ph.D. dissertation "Improving the Signal-to-Noise Ratio by Cross Correlation in Flow Injection Analysis and High Performance Liquid Chromatography" (1990) revealed a method for reducing the signal to noise ratio in high performance liquid chromatography. His approach was to produce a clean (relatively noise free) signal in an artificial setting with high concentrations of the solutes. This clean signal was then cross correlated with the noisy signals produced in the usual settings. Although this approach was successful, it brings up the question of how to produce a clean signal when the solutes are unknown. Also, McKean used chromatography equipment of the late 1980's.

McKean discusses the ensemble averaging of multiple signals to improve the signal to noise ratio. He does not, however, suggest the use of multiple sensors and indicates the averaging approach would be "time consuming" for high performance liquid chromatography, presumably due to needing to run multiple identical samples past a single sensor. McKean, because his research focused on his cross correlation method, was not motivated to utilize multiple sensors for summing or averaging signals in a complete chromatogram.

Another novel approach was suggested by Hjerten in U.S. 5,114,551 in which a single detector was used to pick up a signal at multiple locations on the capillary. This was done by looping the capillary around and returning back to the light source and sensor location. A relatively small improvement in the signal to noise ratio would be realized with this method due to the limited number of chromatograms that can feasibly be taken. In one embodiment of this invention, the capillary tube is moved laterally in order to move a new portion of the capillary tube between the light source and sensor. This is an unnecessary complication, requiring control circuitry

and moving parts that can fail. The flow, too, is not favorably enhanced by looping or by moving the capillary tube.

There is, therefore, a need for a way to significantly improve the signal to noise ratio of the signal produced by photocells in capillary high-pressure liquid chromatography with no moving parts.

Summary of the Invention

A purpose of this invention is to provide a method and device capable of producing a substantially clean signal in high-pressure liquid chromatography. Another purpose of the present invention is to carry out the aforementioned purpose with no moving parts. As part of these purposes, an objective of the present invention is to accurately track a solvent spike as it flows along its capillary tube.

It is well known that the sensitivity of a chromatograph, produced by any of the capillary separation schemes, HPLC, CE, or, CEC, is directly proportional to the path length of a beam of light passing through the sample. An indirect method of increasing this path length is to repeatedly take a chromatograph of the same particle of solution. A linear array of monochromatic light sources and sensor photocells are aligned parallel with a polished quartz capillary tube. Some of the first photocells encountered by the solvent spike are used to accurately locate the solvent spike, positively identifying a fluid particle. The spike is a relatively large fluid particle about the same size as the injection volume. It arises when the injecting solvent has a refractive index differing from that of the eluting solvent at the same time and location as the injection. Because the scanning tube is a cylinder, the light seen by each photocell varies sharply as the solvent spike (refractive index spike) passes them by. It differs from the absorbance (chromatographic) signal in that the solvent spike is usually taller and arrives first. Therefore it is easy to detect as it is located or goes past any photocell location. In applications involving CE or CEC, and sometimes also capillary HPLC, the solvent spike may be replaced by an unretained, or otherwise leading UV absorbing reference peak. This is accomplished by including a selected absorbance reference material added to the sample. In a first embodiment of

the invention, the flow of the solution is stopped within the capillary tube with the solvent spike oriented at a known photocell. The photocell sensors in the linear array are used to scan the solution to obtain a chromatogram for the particular sample. Each photocell will either scan repeatedly, and the scans summed, averaged, or statistically correlated, or the scan will be taken over time and integrated.

In a second embodiment of the invention, the solution flows as usual and a solvent spike in the solution is accurately tracked as it is scanned by a series of photocells in the linear array. Signals from each of the photocells, as the same particle of fluid passes through the associated light beam, are summed or statistically correlated. Because the same particle of solution is being tested, the pertinent information in the signal is the same for each reading. The noise should not be correlated to this method of taking multiple readings. The resulting sum (or average, or statistical correlation) has an improved signal to noise ratio because the signal is strengthened by a factor of N (where N is the number of photocell sensors used to scan a particular fluid particle), while the noise is only strengthened by a factor of $\sqrt{N}$ (assuming white noise).

A third embodiment of the invention is described as follows. The chromatogram signal may be divided up into segments, each segment having a predetermined size based on fluid volume or flow. A total number of groups is ten or more. Within each of the segments, the signal segments are summed, integrated, or statistically correlated. Then the resulting signal or values for all segments are concatenated to produce a total signal having less noise than the original. This embodiment is particularly adapted for use with a sequential separation technique, such as a chromatogram. However, the present embodiment is not limited to such a sequential separation technique, nor even the use of a plurality of sensors. Any analog or digital signal may be operated on as specified, above. The segments of this third embodiment may or may not overlap.

The novel features believed to be characteristic of this invention, both as to its organization and method of operation together with its further objectives and advantages, will be better understood from the following description considered in

connection with the accompanying drawings in which a presently preferred embodiment of the invention is illustrated by way of example. It is to be expressly understood however, that the drawings are for the purpose of illustration and description only and not intended as a definition of the limits of the invention.

Brief Description of the Drawings

Fig. 1 shows a light source, capillary tube and photocell sensors.

Fig. 2 shows a flow chart showing steps for carrying out the present invention with a quiescent solution and taking a continuous reading.

Fig. 3 shows a flow chart showing steps for carrying out the present invention with a quiescent solution and taking multiple readings.

Fig. 4 shows how information is carried from an array of photocell sensors, ultimately to a chromatogram when the solution is quiescent during scanning.

Fig. 5 shows a representative clean signal produced by a linear array of 1024 photocell sensors.

Fig. 6 shows a representative noisy signal produced by a linear array of 1024 photocell sensors.

Fig. 7 shows a cleaned signal produced using the methods of the present invention with a linear array of 1024 photocell sensors.

Fig. 8 is a flow chart showing steps for carrying out the present invention with a flowing solution.

Fig. 9 shows how information is carried from an array of photocell sensors, ultimately to a chromatogram when the solution is flowing during scanning.

Fig. 10 shows a representative clean signal produced over time by a photocell sensor.

Fig. 11 shows a representative noisy signal produced over time by a photocell sensor.

Fig. 12 shows a representative cleaned signal produced over time by a photocell sensor.

Fig. 13 shows a noisy signal subdivided into sections.

Fig. 14 shows a piecewise averaged signal.

Fig. 15 shows a process going from a noisy signal subdivided into sections to a

piecewise averaged signal.

Fig. 16 shows a process diagram for converting a noisy signal to a cleaner signal.

Fig. 17 shows a noisy signal being denoised by a running average, integration, or statistical correlation.

Best Mode for Carrying Out the Invention

In **Fig. 1** a schematic of an apparatus for the present invention is depicted. A uniform, monochromatic light source (or sources) **100** lines one side of a polished, quartz capillary tube **110**. Directly opposite (on the other side of the capillary tube **110**) is a linear array of photocell sensors **120**. (Only twelve individual photocell sensors **141–152** are shown in **Fig. 1**, however, in practice many more individual photocell sensors **141–152** would be used.) Some of the light emitted from the light source **100** is reflected off the capillary tube **110** and the solution flowing through the capillary tube. Some of the light is absorbed by the solution. That light not reflected or absorbed, passes through the capillary tube **110** and the solution flowing in the capillary tube. Each of the photocell sensors **141–152** creates a signal related to the light intensity of the light that impinges on it. An identifying feature of the components of the solution is the amount of light absorbed.

A signal of interest is one over a period of time. A fluid particle **130** is defined as a small mass of fluid of fixed identity. As a fluid particle **130** of the solution is scanned by a photocell **141–152**, a signal is recorded based on the light passing through the fluid particle **130** and impinging on the photocell sensor **141–152**.

In the first embodiment of the present invention, the fluid is quiescent when the scanning step is carried out. Therefore, any given photocell **141–152** records data for a single fluid particle **130**. The data recorded might be an integral of each individual photocell's **141–152** signal over time, or a summation, average or statistical correlation of a series of the photocell's **141–152** signals taken sequentially over time.

In the second embodiment of the present invention, the same particle of fluid **130** travels past each of the photocell sensors **141–152** in turn. The velocity of the fluid particle **130** is determined as follows. A location of a solvent spike is detected using

the first photocell sensors **141–152** the solvent spike encounters. For instance, 1000 photocell sensors **141–152** in an array **120** of 2048 photocell sensors may be used to detect a solvent spike. It is known that the solvent spike is a fluid particle about the same size as the injection volume. The solvent spike is formed when the injecting solvent has a refractive index differing from that of the eluting solvent at the same time and location as the injection. The light detected by each photocell sensor **141–152** varies sharply as the solvent spike (refractive index spike) passes between the light source **100** and a photocell sensor **141–152**. The signal produced by a photocell sensor **141–152** when the solvent spike is scanned differs from the absorbance (chromatographic) signal in that the signal due to the solvent spike is usually taller and arrives before other peaks. The said spike may also be an absorbance spike. Once the spike has been detected, its velocity is determined using a small number of additional photocell sensors **141–152**, for instance, as the spike passes from the 1000^{th} photocell sensor to the 1005^{th} photocell sensor. This velocity, in photocell sensors per unit time, is multiplied by a predetermined factor, for example 1001, to obtain a scanning speed, again in photocells per unit time.

Using this scanning speed, a number of photocell sensors **141–152** equal to the predetermined factor (e.g. 1001) are scanned. Using the example, above, the 6^{th} through the 1006^{th} photocell sensors are scanned. At the end of this scan, the solvent spike is again located using the data just obtained from the photocell sensor scan. The solvent spike should be at the last photocell sensor from which data were obtained. According to our example, the solvent spike should be located at the 1006^{th} photocell sensor. If the solvent spike is not located at the correct sensor, the solvent spike velocity is recalculated from the new data, a new scanning velocity is calculated and the process repeated using the correct bank of photocell sensors **141–152**. In our example, this new correct bank of photocell sensors **141–152** for the next step would be the 7^{th} through the 1007^{th} photocell sensors.

Figs. 2 and **3** depict flow diagrams for the first embodiment wherein the solute is quiescent when scanned. Initially, the solute is pumped into the capillary tube **110**. The location of the solvent spike is detected **200** as the solute flows in the same

manner as described above. When the solvent spike arrives at a predetermined photocell sensor **141–152**, for example the 1000[th] in an array **120** of 1024 photocell sensors, the solute flow is stopped **210**, either by stopping the pump or by closing a valve. With the solute in a quiescent state, and the solvent spike located at a known photocell sensor **141–152**, all the photocell sensors **141–152** in the array **120**, or a predetermined subset of the photocell array **120**, are scanned continuously **220** (**Fig. 2**) for a predetermined duration. The result of this continuous scan is integrated **230** in the photocell sensor **141–152** if its charge storage capacity is adequate, or in a separate integrator for each photocell sensor **141–152**. The resulting information is a plot of the light passing through the capillary tube **110** with the independent variable (abscissa) being the photocell number from which the signal was taken (see **Fig. 7**). This abscissa can easily be converted to the location along the capillary tube **110**, x, if preferred.

In **Fig. 3**, the same approach as that shown in **Fig. 2** is taken, except that instead of a continuous scan **220**, scans are taken repeatedly **320** and then summed, averaged, or statistically correlated **330** to produce the reduced-noise signal.

The process of scanning, calculating, and storing the data is depicted in **Fig. 4**. The linear photocell sensor array **120** is shown at the top with twelve photocell sensors **141–152** shown. The ellipses shown to the left of photocell sensor **141** and to the right of photocell sensor **152** indicate there may be more photocell sensors.

The signal from each of the photocell sensors is fed into a set of processing blocks **440–453**. These processing blocks **440–453** may include an Analog to Digital (A/D) converter and an integration, summation, or statistical correlation function. The processing blocks **440–453** may be inherent to the photocells, themselves, if the charge storage capacity of each photocell sensor is adequate for the task, or they may be separate units, carrying out their operations in analog or digital mode. Finally, the resulting, processed data are organized into a chromatogram, as indicated by the plot **470** shown.

A representative plot of a scan is shown in **Fig. 5**. This plot shows a noise-free signal of the quiescent solution as taken from 1024 photocells. Here, four peaks or

spikes are shown. On the abscissa is the photocell sensor number from 1 to 1024, while the ordinate is the signal, as amplified from the photovoltaic sensors, in volts.

In **Fig. 6** the noise-free signal is shown with white noise superimposed upon it, resulting in a noisy signal. The white noise has a maximum amplitude of three volts. The clean signal cannot be identified due to the noise.

In **Fig. 7**, 100 noisy signals like that shown in **Fig. 6** have been integrated with the time of integration divided out, or arithmetically averaged, or statistically correlated. As can easily be seen, due to the averaging step, the relative fraction of the signal attributable to noise is greatly reduced.

A flow diagram of the second embodiment of the present invention is shown in **Fig. 8**. Once the solvent spike is detected **200**, its speed, V_{ss}, is estimated **800** at photocell sensor *m*, such as the 1005th photocell sensor, as used in the example, above. Then the *n* ($n = 1001$ in the above example) photocell sensors from $m+1$ "back" (upstream) are scanned at a rate, V_{sr}, such that, when the scan is finished, the solvent spike should have reached photocell sensor $m+1$ **810**. The scanning speed is calculated as $V_{sr} = nV_{ss}$, where both V_{sr} and V_{ss} are in photocells per unit time.

After the aforementioned scan **810**, the location of the solvent spike is, again, detected **820**, ultimately to ascertain that it did, in fact, reach photocell sensor $m+1$, and no further. Before comparing the location of the solvent spike to photocell sensor $m+1$, the value of *m* is incremented up by one (1.0) **830** and this new value of *m* is tested **840** against *N*, the total number of photocell sensors **141–152** (2048 in the example above), so the process ends when the last photocell sensor is encountered. If $m \leq N$ at this point, the location of the solvent spike is compared **850** with the location of the photocell sensor *m*. If the solvent spike is at photocell sensor *m*, the same estimated solvent spike speed is used and the process repeated, scanning *n* photocell sensors upstream from and including the new photocell sensor $m+1$ **810**. If the solvent spike is not at photocell sensor *m*, a new solvent spike speed is estimated **800** before the remainder of the process is repeated as before.

For the present embodiment wherein measurements are made of the flowing solution, the photocell sensors **141–152** are again shown in **Fig. 9** with ellipses shown

at each end of the linear array **120** to indicate there may be more photocell sensors than shown. The analog signals from the photocell sensors **141–152** are converted to digital signals in an A/D converter **900**. Because the front of the *n* photocell sensors is shifted such that it moves with the flow, and only *n* photocell sensors are read at each scan, the digital signals, from the first to the last, need only to be stored in memory locations **941–952** from the first to the last. No more shifting is required.

From the memory locations **941–952**, the signal is processed in a calculation function **910** that integrates, sums or statistically correlates each photocell sensors' **141–152** signal to produce a chromatogram, as indicated by the plot **970**.

A chromatogram, such as would be produced by the second embodiment of the present invention, is shown in **Fig. 10**. **Fig. 10** represents a clean (noiseless) HPLC with four peaks. Although the abscissa could be the photocell numbers of the *n* photocell sensors used for each scan, it is just as logical to make the abscissa be time in seconds. The ordinate is, again, the signal, as amplified from the photovoltaic sensors, in volts.

The same signal as shown in **Fig. 10** is replotted in **Fig. 11** with simulated noise superimposed on the clean signal. The white noise has a maximum amplitude of three volts. The clean signal cannot be identified due to the noise.

The next step in the analysis is shown in **Fig. 12**. Here, 100 noisy signals, with the same clean content as shown in **Fig. 10** and different noise (all with a maximum amplitude of three volts), were averaged. The improvement can easily be seen when comparing **Fig. 12** with **Fig. 11**. The improvement is evident, even though only 100 readings were used (in practice, many more could be used). Even the last and smallest peak (seen in **Fig. 10** at about 515 seconds) can be resolved from the noise. The signal could also have been summed or statistically correlated.

A third embodiment of the present invention is shown in **Figs. 13–17**. In **Fig. 13,** a noisy signal is shown. The dashed lines indicate how the noisy signal is subdivided into segments **1300**. The signal within each segment is averaged, integrated, or statistically correlated. The results are recombined into a new, less noisy signal such as shown in **Fig. 14**. The process is illustrated, symbolically, in

Fig. 15, where the arrows indicate the flow of information from the noisy signal at the top to the piecewise average, integrated, or statistically correlated signal at the bottom.

A flow diagram of this third embodiment is shown in **Fig. 16**. A noisy signal **1600**, which may be an analog signal or a digital signal, from a single sensor or a plurality of sensors, is divided into a plurality of segments by a signal divider **1610**. There are ten such segments indicated in **Fig. 16**, but this should be considered a minimum. Each segment passes through a separate operation **1620**, which could be summation, integration or statistical correlation. The resulting signals are concatenated in a concatenation function **1630**, to recreate the signal with less noise. The concatenated signal may, optionally, be passed through a filter **1640**. A suitable filter **1640** is a low-pass filter. An analog filter is used for an analog signal, while a digital signal is used for a discrete signal. The filter **1640** will have the characteristic that it attenuates noise corresponding, in time or frequency, to noise associated with time or frequency of one of the signal segments from the signal divider **1610**. The result, regardless of the use of the optional filter, is a signal **1650** that is cleaner (less noisy) than the noisy signal **1600**.

A variation on this third embodiment utilizes a *running* average or other operation wherein each point of a first, noisy signal is replaced by its average, integral, or statistical correlation over a neighborhood of the original point to produce a cleaner signal. This approach is also valid for discrete or analog signals. **Fig. 17** depicts the present embodiment for a discrete signal. In **Fig. 17**, only three points are shown being used for the operation **1620** which, again, could be summation (or averaging), integration, or statistical correlation. In reality, many more points would be used. The top signal depicts a noisy signal. Each point of the lower, cleaner signal is a result of the operation **1620** of a number of points or regions of the noisy signal. Each point of the lower, cleaner signal contains information from some points of the noisy signal also used to determine points neighboring the clean signal point. This variation is not limited to a discrete signal.

Obviously many modifications and variations of the present invention are

possible in light of the above teachings. Any number of photocell sensors may be used in the linear arrays. The photocell sensor array need not be linear. It is, therefore, to be understood that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described.